Journal of Earth Science and Engineering 3 (2013) 371-378

Approaches to Detecting Gene-Environment Interactions in Human Variation Using Genetic Engineering, Remote Sensing and GIS

Sumiko Anno[1], Kazuhiko Ohshima[2], Takashi Abe[3], Takeo Tadono[4], Aya Yamamoto[5] and Tamotsu Igarashi[5]

1. Department of Civil Engineering, Shibaura Institute of Technology, Tokyo 135-8548, Japan

2. Department of Computer Bioscience, Nagahama Institute of Bio-Science and Technology, Shiga 526-0829, Japan

3. Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan

4. Earth Observation Research Center, Japan Aerospace Exploration Agency, Ibaraki 305-8505, Japan

5. Research and Development Department, Remote Sensing Technology Center of Japan, Tokyo 105-0001, Japan

Received: April 10, 2013/Accepted: May 15, 2013/Published: June 15, 2013.

Abstract: Elucidation of the relationships between genetic polymorphisms and environmental exposures can provide insights into the pathways and mechanisms underlying complex traits. A new approach was used to detect G×E (gene-environment) interactions involved in human skin pigmentation variation to better understand the adaptive evolution of skin pigmentation. Specifically, we used genetic engineering, remote UVR (ultraviolet radiation) sensing and GIS (geographic information systems) to integrate the analysis of genetic and environmental factors into a coherent biological framework. Since we expected to generate large datasets for this multidimensional analysis, we used PCA (principal components analysis) as a spatial statistical analysis technique for analyzing the G×E interactions. The results suggest that skin pigmentation may be affected by mutations induced by UVR and support the hypothesis that global variation in skin pigmentation may be the result of localized adaptation to different UVR conditions via natural selection. Analyzing the relationships between heterozygous frequencies for SNP (single nucleotide polymorphism) loci and seasonal UVR levels as the environment changes will help elucidate the selective mechanisms involved in the UVR-induced evolution of skin pigmentation. Skin pigmentation fulfills the criteria for a successful evolutionary G×E interactions model.

Key words: RS (remote sensing), GIS, genetic engineering, G×E interactions, adaptability.

1. Introduction

By definition, a phenotype results from interactions between genes and the environment [1]. Understanding the role of adaptive evolution in skin pigmentation requires an approach to G×E interactions that integrates results of the analysis of genetic and environmental factors into a coherent biological framework. Elucidation of the relationships between genetic polymorphisms and environmental exposures can provide insights into the pathways and mechanisms underlying complex traits [2].

Studies of G×E interactions in diseases have successfully used several approaches involving different methods. Such studies have identified high-risk subgroups in the population [2], found novel potential risk-factor combinations [3], provided insights into pathway mechanisms for complex diseases [2] and maximized the use of available biomedical data to improve our understanding of complex diseases [4]. Although, these methods have been applied to diseases that are thought to result from the combined effects of genes, environmental factors and G×E interactions, these methods have not been exploited to elucidate the mechanisms underlying complex human phenotypes

Corresponding author: Sumiko Anno, associate professor, research fields: RS, GIS, application of RS and GIS technologies in public health. E-mail: annou@sic.shibaura-it.ac.jp.

such as skin color variation.

There are quantitative data that indicate strong correlations between skin reflectance and UVR (ultraviolet radiation) levels, supporting the hypothesis that skin color variation is due to the effects of genes, environmental factors and G×E interactions. For example, the distribution of skin color variations in indigenous peoples is quantitatively related to remotely-sensed UVR measurements, demonstrating that skin reflectance correlates strongly with absolute latitude and UVR levels [5]. This work further suggested that global variations in skin pigmentation might be the result of localized adaptation to different UVR conditions via natural selection [5, 6]. Moreover, human skin pigmentation was shown to be the product of two clines produced by natural selection that adjusted the levels of constitutive pigmentation to UVR levels. One cline was generated by high UVR near the equator and led to the evolution of dark, photoprotective and eumelanin-rich pigmentation. The other cline was produced by the requirement for UVB (ultraviolet B) photons to sustain cutaneous photosynthesis of vitamin D_3 in low-UVB environments, resulting in the evolution of depigmented skin [7].

Despite the importance of G×E interactions for complex phenotypes, there has been little progress in developing methods that can detect and clarify the interactions involved in skin color variation. Here we used a spatial-based statistical approach that provides a useful and powerful theoretical framework for investigating the evolutionary and adaptation mechanisms involved in skin color variation by detecting G×E interactions using a combination of genetic engineering, RS (remote sensing) and GIS (geographic information systems).

2. New Approaches

Fig. 1 shows a conceptual framework for using new approaches to detect G×E interactions which are important for human skin pigmentation variation using genetic engineering, RS and GIS. In this study, genetic engineering was used as a molecular biological approach, RS and GIS were used as an ecological approach to elucidate the relationships between genetic characteristic and UVR exposure.

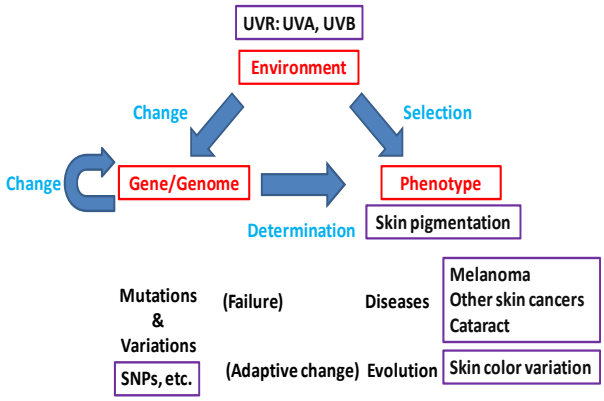

Fig. 1 The conceptual framework of two approaches to analyzing the gene-environment interactions that are important for skin color variation.

3. Molecular Biological Approach

We collected data for 20 SNPs (single nucleotide polymorphisms) (rs819136, rs1129414, rs2075508, rs10960756, rs3793976, rs2298458, rs3212363, rs1805008, rs3212371, rs2279727, rs4778182, rs1800419, rs2311843, rs1800414, rs1800404, rs7623610, rs704246, rs16964944, rs1724577 and rs4776053) in seven candidate genes (ASIP,TYRP1, TYR, MC1R, OCA2, MITF and MYO5A) involved in human skin pigmentation. It was found SNP alleles at multiple loci that could be considered haplotypes that contribute to substantial differences in skin pigmentation in Europeans versus East Asians. The haplotype-allele combination rs2311843-C, rs1800404-A and rs4776053-C was associated with the European group and found in the genes for OCA2 and MYO5A on chromosome 15. The haplotype-allele combination rs1800419-C/rs1800414-G/rs1800404-G was associated with the East Asian group and was found in the gene for OCA2 on chromosome 15 [8].

We also detected natural selection in candidate pigmentation genes in haplotypes that were revealed by SNP analyses. We analyzed the

rs2311843/rs1800404/rs4776053 haplotype in the OCA2 and MYO5A genes on chromosome 15 in the European population group and the rs1800419, rs1800414, rs1800404 haplotype in the OCA2 gene on chromosome 15 in the East Asian population group. We conducted both Tajima's *D* test and Fu and Li's *F* test using DnaSPVersion 5 to determine whether any of the SNPs were under natural selection pressures. Only data that fit the Hardy-Weinberg equilibrium were used in the analysis. The Tajima's D value of the rs1800419/rs1800414/rs1800404 haplotype in the East Asian population was significantly positive ($D = 2.83967$, $P < 0.01$). However, using Fu and Li's *F* test, the value of the haplotype was not significant ($F = 1.70954$, $0.05 < P < 0.10$). No significant characteristics were detected for the other haplotypes. These results indicate the possibility that the haplotype in the OCA2 gene in the East Asians population has been under selective pressure [9].

4. Ecological Approach

The UVR data were derived from readings taken from the NASA (National Aeronautics and Space Administration) and TOMS (Total Ozone Mapping Spectrometer), which was flown aboard the Nimbus-7/Earth Probe satellites. The TOMS sampled single wavelengths representative of long-wave and medium-wave UVR: 324-nm and 380-nm wavelengths for UVA (ultraviolet A) (range, 315-400 nm) and 305-nm and 310-nm wavelengths for UVB (range 280-315 nm). The original data set was very large, comprising over 64,800 readings taken every day from 1979 to 2003. Abridged data sets were produced for each of the four wavelengths using the average reading for each month from 1979 to 2003. The abridged data set for 310-nm UVB (which induces deamination and causes a barely perceptible reddening of light skin) was then used along with the points of longitude and latitude into a GIS for spatial analysis and interpolated using IDW (inverse distance weighting) for the values of a raster. The raster layer was overlaid with the polygon layer, which contained three polygons representing the birthplaces of the human race (Africa, European and East Asia) [10] (Fig. 2). Each polygon contained and summarized the raster values within its area and reported the results as a table for spatial statistical analysis. Chaplin's study [6] found that the evolution of skin reflectance could be almost fully modeled as a linear effect of UVR in fall alone. The monthly mean values by the human racial population were organized for the seasonal mean values for winter, spring, summer and fall.

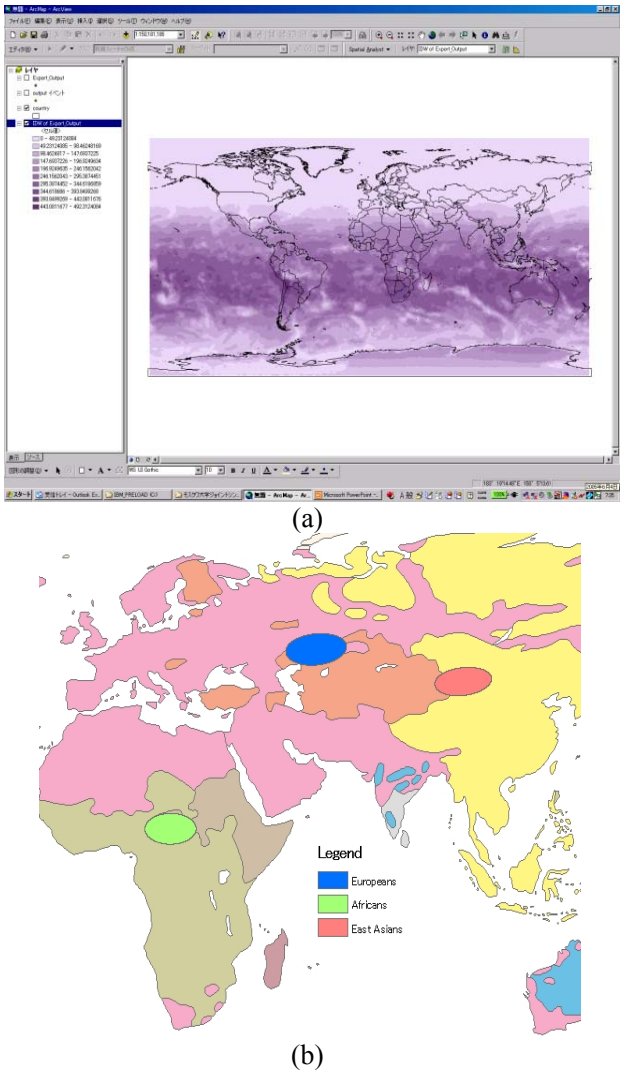

Fig. 2 Spatial analysis in geographic information systems: After the 310-nm UVB raster data (a) were overlaid with the polygon layer in which the three polygons represent the birthplaces of the human race (b), the three polygons contained the mean monthly 310-nm UVB value.

5. G×E Interaction Analysis

With the data obtained from these analyses based on genetic engineering, RS and GIS data, we then used spatial statistical analysis to evaluate G×E interactions. We assessed the relationship between the SNPs at candidate genes involved in human skin pigmentation and seasonal UVR exposure in the three regions denoted using polygons. This analysis was intended to clarify the mechanism underlying the molecular basis for changes in the genetic background of human skin color variation and for human adaptability from the perspective of human evolution.

To perform these analyses, we used genotype frequency data for 20 SNPs in seven candidate genes for human skin pigmentation that were obtained in our previous study plus the data for 553 SNPs in the OCA2 gene (a common gene in both European and East Asian populations in our previous study). We then calculated the mean values for the heterozygous genotype frequency for every allele pattern in each population.

Next, we conducted PCA (principal component analysis) using the correlation matrix to determine the relationships between SNP frequencies for heterozygous subjects and the seasonal UVR data. The following four kinds of PCA analyses were conducted: (I) 20 SNPs in seven candidate genes from our previous study + UVR data for the fall; (II) 20 SNPs in seven candidate genes from our previous study + UVR data for all four seasons; (III) 553 SNPs in the OCA2 gene from the HapMap database + UVR data for the fall; (IV) 553 SNPs in the OCA2 gene from the HapMap database + UVR data for all four seasons.

6. Results

Table 1 shows the mean SNP frequencies of the indicated heterozygous alleles in three populations that differ in terms of typical skin pigmentation. This analysis was performed using both the SNPs obtained from our previous study and those obtained from the HapMap database. The results showed more significant correlations in the mean G/T SNP frequencies for heterozygous alleles for the SNPs identified in our previous study than in those from the HapMap database.

The correlation matrix results showed very high correlations between the G/T SNP and the meanseasonal 310-nm UVR levels (Table 2). The mean 310-nm UVR levels in fall and winter were more highly correlated with the G/T SNP than the mean seasonal 310-nm UVR levels in spring and summer. PCA were performed on both the SNPs obtained from our previous study and on those from the HapMap database, and the correlations were higher for

Table 1 Mean SNP frequencies for the heterozygous alleles.

Environment	UVR (310 nm, Autumn)	123.523	26.708	51.715
	20 SNPs in 7 genes	Africans	Europeans	East Asians
Genome	SNP (C/T)	0.247	0.345	0.323
Genome	SNP (A/G)	0.206	0.208	0.399
Genome	SNP (G/T)	0.405	0.093	0.131
Genome	SNP (A/T)	0.545	0.481	0.413
Genome	SNP (A/C)	0.367	0.100	0.500
	553 SNPs in *OCA2* gene	Africans	Europeans	East Asians
Genome	SNP (A/C)	0.337	0.299	0.180
Genome	SNP (A/G)	0.260	0.234	0.187
Genome	SNP (A/T)	0.199	0.271	0.157
Genome	SNP (C/G)	0.292	0.246	0.153
Genome	SNP (C/T)	0.279	0.241	0.208
Genome	SNP (G/T)	0.251	0.155	0.137

Table 2 The correlation matrices for cases I and II.

	UVR (Autumn)	UVR (Summer)	UVR (Winter)	UVR (Spring)
UVR (Autumn)	1.00000	0.90878	0.99078	0.99573
UVR (Summer)	0.90878	1.00000	0.84387	0.94341
UVR (Winter)	0.99078	0.84387	1.00000	0.97405
UVR (Spring)	0.99573	0.94341	0.97405	1.00000
SNP (C/T)	−0.99905	−0.88973	−0.99574	−0.99076
SNP (A/G)	−0.27792	0.14828	−0.40549	−0.18807
SNP (G/T)	0.99023	0.84172	0.99999	0.97314
SNP (A/T)	0.70147	0.34009	0.79156	0.63270
SNP (A/C)	0.42726	0.76556	0.30084	0.50888

the SNPs from our previous study. In contrast to the G/T SNP, the major UVR variables were negatively correlated with both the C/T SNP and the A/G SNP variables in cases I and II such that an increase in the major UVR variable(s) corresponded to decreases in both the C/T SNP and the A/G SNP variables.

The biplots of the PCA showed the principal component scores of each population (Africans, Europeans and East Asians) represented as points on the first two principal components axes. The points of the principal component scores of each distinct population were distributed separately, dividing the data into two groups: one group of Africans and another group of Europeans and East Asians for the cases I, II and IV.

The principal component loading results showed a high first principal component loading of 0.98 for the G/T SNP, which was positively correlated to the G/T SNP variable in case I. The other principal component loadings were low for the other SNPs and did not correlate with other SNPs for cases II, III and IV. This indicates a strong relationship between the G/T SNP and fall.

7. Discussion

A multidimensional analysis like this one is expected to involve a large dataset; therefore, our method included PCA. Skin pigmentation is determined by the DNA in somatic cells and genetic mutations are passed on to the next generation by germ-line DNA. If mutations in skin pigmentation genes were passed on to the next generation, they could lead to changes in skin pigmentation over time. Our results suggest the possibility that genes involved in skin pigmentation might be subject to UVR-induced mutations. Chaplin's study [6] found that the evolution of skin reflectance could be modeled almost fully as a linear effect of UVR in fall aloneand our results also indicated a strong relationship between the G/T SNP and fall.

We identified significant correlations between the mean G/T SNP frequencies for heterozygous alleles and all mean seasonal UVR values for the three populations. The biplots from the PCA analyses showed that the subjects were divided into two groups: one group of Africans and another group of Europeans and East Asians. The first principal component scores were positive for Africans and negative for Europeans and East Asians in case I, while the first principal component scores were negative for Africans and positive for Europeans and East Asians in case II. Furthermore, all PCA analysis results showed strong positive correlations between the major seasonal UVR variable(s) and the G/T SNP compared to other SNPs. Stronger correlations were found for the data set obtained with the SNPs identified in our previous study. These findings suggest that the G/T SNP and the seasonal 310-nm UVR contribute to the separation of Africans and Europeans/East Asians.

Zhang's research on the DNA mutation patterns of human ribosomal protein pseudogene sequences revealed that nucleotide transitions (C:G→T:A or T:A →C:G) are more common than transversions [11]. Of

the two transitional events, C:G→T:A occurs much more frequently than T:A→C:G [11]. The former transition can be caused by a change in cytosine methylation and has a larger substitution rate than other base substitutions. Two transversional events, i.e., G → T and T → G, neither occurred more frequently than the other mutations. However, our analyses found very high correlations between the G/T SNP frequencies in genes for skin pigmentation with the mean seasonal 310-nm UVR. It is possible that the whole genome does not have significant characteristics, but, the genes for skin pigmentation may have such a tendency.

Available data also indicate that mutations possibly occurred due to modification of bases, i.e., 8-oxoG (8-oxoguanine). Among the different types of DNA damage, oxidative DNA lesions caused by ROS (reactive oxygen species)—Which are generated both as a byproduct of oxidative metabolism and as a consequence of exposure to ionizing radiation and other environmental factors—are considered a major threat to the genome [12, 13]. Of the four bases, guanine is the most susceptible to oxidation, and its simple oxidized form, 8-oxoG, is a major product when DNA or nucleotides are subjected to oxidation [14]. 8-oxoG is a potent pre-mutagenic lesion because it can pair with adenine as well as cytosine during DNA replication and can thus cause a G:C→T:A transversion mutation [15]. It is thus possible that 8-oxoG formation, which is involved in active oxygen through UV exposure, may contribute to G:C→T:A transversion mutations. The mutations may then be passed on to the next generation, leading to skin pigmentation mutations and resulting in skin pigmentation variations. This supports the theory that depigmented and tannable skin may have evolved numerous times in hominin evolution via independent genetic pathways under natural selection.

8. Conclusions

Analyzing the relationships between heterozygous frequencies for SNP loci and seasonal UVR levels as the environment changes will help elucidate the selective mechanisms involved in the UVR-induced evolution of skin pigmentation. Skin pigmentation fulfills the criteria for a successful evolutionary G×E interactions model. First, skin pigmentation was produced by an imperfect replicator [7]. That is, skin pigmentation is transmitted to the next generation by germ-line DNA, which is possibly subject to mutation induced by UVR. Second, natural selection occurs through differential survival and reproduction rates of different phenotypes exposed to different UVR levels. Furthermore, skin pigmentation is an attractive model system for understanding and teaching evolutionary G×E interactions because it is readily visible and the basic contributory mechanisms are easily understood.

Europeans who are born and reach adulthood in Europe and then migrate from Europe to Australia appear to have one-fourth of the risk of developing melanoma compared to those of European descent who are born and brought up in Australia. On the other hand, Europeans who are born in Europe and migrate from Europe to Australia during childhood have melanoma development rates that are similar to those of Europeans who were born and brought up in Australia. This may indicate that exposure to intense UVR in childhood and adolescence causes melanoma development [16]. It may further suggest that Europeans who are born and who live in regions at high altitudes have a reduced ability to adapt to the strong UVR that is present at low latitudes. Therefore, it is important to investigate human adaptability to UVR exposure to enable better prediction of the health risks caused by extreme environmental conditions and to develop preventive interventions.

Despite the importance of G×E interactions in complex phenotypes, there has been little progress in developing methods that can detect the interactions involved in skin color variation in human populations. In this study, we used a spatial statistical approach that provides a useful and powerful theoretical

framework for investigating the adaptation mechanisms involved in skin color variation by detecting G×E interactions. The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data-mining techniques for extracting spatial patterns. Spatial statistical analysis enables the discovery of interesting, previously unknown, potentially useful patterns from spatial databases. Efficient tools for extracting information from geospatial data are crucial to organizations that make decisions based on spatial datasets and are necessary in a variety of fields, including ecology and environmental management, public safety, transportation, earth science, epidemiology and climatology [17, 18].

Several factors must be considered when trying to clarify the evolutionary adaptations of skin color variation that result from localized adaptation to UVR conditions through natural selection. First, information on genetic and environmental factors must be integrated into a single database. Second, the evolution of the genome occurs through the accumulation of environmentally-associated rules that correspond to the cumulative knowledge the genome has acquired regarding its environment [1]. Third, when UVR is the influential environmental factor, GIS can be used to analyze and visualize spatial patterns. Finally, the identification of functional relationships between molecular evolutionary genetics (e.g., heterozygous frequencies for SNP loci) and seasonal UVR levels requires computational methods that utilize spatial datasets.

Our PCA-based approach to spatial statistical analysis is a convenient way to use association rules to extract spatial associations between genetic and environmental variables that determine skin color variation in different human populations. This approach may reveal environment-based differences in genetic traits and can maximize the use of available data to improve our understanding of the adaptive mechanisms that give rise to human skin color variations. Spatial statistical analysis may identify biological mechanisms that regulate human adaptability to UVR. Moreover, identifying the complex mechanisms that have shaped evolutionary changes in skin color may advance our understanding of the history of human adaptation to local environments and may potentially have important public health implications. Understanding human adaptability to UVR may allow us to predict the impact of environmental changes on health risks, particularly regarding high UV exposure secondary to ozone depletion.

Traditional parametric statistical approaches have limited power for modeling the high-order, nonlinear interactions that are probably important in generating complex phenotypes. Given the newly developed RS technologies that capture more detailed, higher resolution UV wavelength measurements and the emerging approaches to G×E interactions that integrate the data into a coherent biological framework, it seems likely that the mechanism underlying human adaptability to varying levels of UVR will be clarified over the next several years. Studies of skin pigmentation at all levels, i.e., at the genetic level to the population level, will help provide a systematic understanding of this phenomenon. New powerful spatial statistical tools and automated learning methods will be developed to computationally model the relationships among genetic variables, protein variables, cell variables, tissue variables, organ variables, physical variables, environmental exposure and skin color variation.

References

[1] K.J.L. Irizarry, B. Merriman, M.E. Bahamonde, M.L. Wong, J. Licinio, The evolution of signaling complexity suggests a mechanism for reducing the genomic search space in human association studies, Molecular Psychiatry 10 (2005) 14-26.

[2] C.E. Murcray, J.P. Lewinger, W.J. Gauderman, Gene-environment interaction in genome-wide association studies, American Journal of Epidemiology 169 (2009) 219-226.

[3] E.J. Duell, P.M. Bracci, J.H. Moore, R.D. Burk, K.T. Kelsey, E.A. Holly, Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer, Cancer Epidemiology, Biomarkers & Prevention 17 (2008) 1470-1479.

[4] R.C. McEachin, B.J. Keller, E.F. Saunders, M.G. McInnis, Modeling gene-by-environment interaction in comorbid depression with alcohol use disorders via an integrated bioinformatics approach, BioData Mining 1 (2008) 1-13.

[5] N.G. Jablonski, G. Chaplin, The evolution of human skin coloration, Journal of Human Evolution 39 (2000) 57-106.

[6] G. Chaplin, Geographic distribution of environmental factors influencing human skin coloration, American Journal of Physical Anthropology 125 (2004) 292-302.

[7] N.G. Jablonski, G. Chaplin, Human skin pigmentation as an adaptation to UV radiation, Proceedings of the National Academy of Sciences 107 (2010) 8962-8968.

[8] S. Anno, T. Abe, T. Yamamoto, Interactions between SNP alleles at multiple loci contribute to skin color differences between caucasoid and mongoloid subjects, International Journal of Biological Sciences 4 (2008) 81-86.

[9] S. Anno, K. Ohshima, T. Abe, The possibility of detecting natural selection in pigmentation candidate genes from haplotype structure as revealed by SNP analyses, Journal of Physical Anthropology 16 (2011) 99-102. (in Japanese)

[10] C. Scarre, Past Worlds: The Times Atlas of Archaeology, Hammond Incorporated & Times Books, New Jersey, 1988.

[11] Z. Zhang, M. Gerstein, Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes, Nucleic Acids Research 31 (2003) 5338-5348.

[12] P.C. Hanawalt, Genomic instability: Environmental invasion and the enemies within, Mutation Research 400 (1998) 117-125.

[13] D.E. Barnes, T. Lindahl, Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annual Review of Genetics 38 (2004) 445-476.

[14] H. Kasai, S. Nishimura, Hydroxylation of deoxyguanosine at the C-8 position by ascorbic acid and other reducing agents, Nucleic Acids Research 12 (1984) 2137-2145.

[15] S. Shibutani, M. Takeshita, A.P. Grollman, Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG, Nature 349 (1991) 431-434.

[16] B.K. Armstrong, A. Kricker, Sun exposure causes both nonmelanocytic skin cancer and malignant melanoma, in: Proceedings of the International Symposium "Environmental UV Radiation and Health Effects", Munich, Germany, 1993, 105-113.

[17] J.F. Roddick, M. Spiliopoulou, A bibliography of temporal, spatial and spati-temporal data mining research, SIGKDD Explorations Newsletter 1 (1999) 34-38.

[18] S. Shekhar, S. Chawla, Spatial Databases: A Tour, Prentice Hall, New Jersey, 2003.